AI Ethics in Education: A contextual model for global south institutions

Carlos Henryque Pompeu Gomes
Instituto Atlântico
Fortaleza, Brazil
carlos_pompeu@atlantico.com.br

Francisco George Costa Torres
Instituto Atlântico
Fortaleza, Brazil
george_torres@atlantico.com.br

ABSTRACT

This study proposes a contextual model for ethical AI governance in educational institutions, addressing gaps in frameworks predominantly developed in Global North contexts. Based on a narrative review of 23 studies, we identify six recurring dimensions: Evaluation, Transparency, Human centrality, Inclusion, Compliance, and Sustainability. The model operationalizes these dimensions through four criteria adapted to Brazilian realities: algorithmic explainability, continuous auditability, plural epistemology, and collective participation.This aligns with Corrêa et al.'s analysis of 200 AI ethics guidelines, which shows 86.5% origination from Global North regions versus only 6% from Global South contexts. Our findings suggest regional variations in ethical priorities, with inclusion and sustainability receiving greater emphasis in Global South frameworks.

Anchored in decolonial epistemology and participatory governance, the framework questions universalizing claims while preserving technical rigor. It lays the groundwork for practical guidance for responsible implementation in educational settings, supports institutional decision-making and policy design, and advances technological sovereignty and epistemic justice across Global South contexts.

Author Keywords

Responsible AI; Ethics frameworks; Decolonial epistemology; Educational institutions; Global South.

ACM Classification Keywords

K.3.1. Computers and Education: Computer Uses in Education; K.4.1. Computers and Society: Public Policy Issues; K.4.2. Computers and Society: Social Issues.

INTRODUCTION

The increasing use of AI in education raises complex ethical questions, particularly following the popularization of generative AI in 2022 [1], [11], [12], [18], [29]. In Latin American contexts such as Brazil, these challenges intensify due to limited infrastructure, structural inequality, and the risk of reproducing Eurocentric logics that render local knowledge invisible [16], [19].

Multilateral organizations such as UNESCO [26], ECLAC [7], and BRICS [5] have developed policy frameworks and proposed global guidelines, while national policy initiatives such as the Brazilian Artificial Intelligence Plan [3] reinforce the centrality of the discussion in the Latin American context. In Brazil, empirical work by Sampaio, Sabbatini, and Limongi [23] proposes guidelines for ethical use of generative AI, emphasizing democratic participation and epistemological diversity.

Following theoretical frameworks from [11] and [16], the literature identifies six fundamental dimensions: evaluation, transparency, inclusion, human centrality, compliance, and sustainability.. We analyzed 23 works published through 2025 to map existing ethical approaches and identify conceptual gaps, positioning university environments as legitimate spaces for algorithmic sovereignty [5], [6].

The main contribution of this work is the proposition of a contextual model for ethical governance of educational AI that integrates technical rigor with decolonial epistemology and participatory governance, offering an operational framework applicable to the Brazilian scenario and other Global South contexts.

Considering this panorama, the present work is guided by the following inquiry: How can ethical AI frameworks be contextualized for Global South educational realities, integrating technical rigor with democratic participation? The working hypothesis is that Global North frameworks are insufficient for the Global South, demanding contextual models that articulate technical, participatory, and epistemological dimensions.

METHODOLOGY

We employed a qualitative approach based on Jobin et al. [15] for mapping emerging fields, complemented by the thematic analysis of Braun and Clarke [4]. This combination is justified by the exploratory nature of the research question and the methodological diversity of the Responsible Artificial Intelligence field in educational contexts, requiring breadth and dialogue between different traditions.

Search Strategy

The search strategy prioritized open databases (SciELO, Google Scholar, institutional repositories from institutions specified below), complemented by public policy and institutional guidelines. The string used was: ("artificial intelligence" OR "AI") AND ("ethics" OR "governance" OR "responsible") AND ("higher education" OR "university") AND ("Brazil" OR "Latin America" OR

"Global South"), adapted according to the database. Searches were conducted through 2025, with SciELO returning 17 specific results, Google Scholar approximately 200 relevant results, and institutional repositories about 50 documents. The initial search identified 267 documents, of which 67 were screened by title and abstract, resulting in 23 studies selected for complete analysis after applying thematic representativeness criteria. The complete table with the studies is available at the following link [30].

Selection and Analysis

Inclusion criteria required focus on educational AI or related topics in higher education from ethical, governance, or social responsibility perspectives. Excluded were purely technical works, studies focused only on basic education, texts without ethical or governance discussion, opinion literature, and documents without peer review. Information about authorship, year, country, institutional context, framework, ethical dimensions, and recommendations was extracted, classifying studies by Global North or Global South based on World Bank criteria.

The analysis followed the methodological framework of Braun and Clarke [4], including familiarization, thematic analysis, and final synthesis. Two independent evaluators conducted complete reading and initial coding of the 23 studies. We resolved disagreements by consensus until consolidating the final result, which increased the procedure's reliability.

The final selection considered three criteria: (1) contribution to the six ethical dimensions (evaluation, transparency, human centrality, inclusion, compliance, and sustainability); (2) foundation of the four operational criteria (explainability, auditability, plural epistemology, and collective participation); and (3) representativeness of the three pillars of the contextual model (algorithmic ethics, decolonial epistemology, and participatory governance).

The selection process followed four stages: (1) initial search in databases with string application; (2) screening by title and abstract applying inclusion/exclusion criteria; (3) complete reading of selected texts; (4) final selection based on contribution to identified ethical dimensions and thematic representativeness. The final distribution included 15 Global North studies and 8 Global South studies, ensuring regional representativeness for comparative analysis.The final distribution included 15 Global North studies and 8 Global South studies, ensuring regional representativeness for comparative analysis. Among the selected works, Brazilian educational institutions analyzed include Universidade Federal do Rio Grande do Norte (UFRN), Universidade de Brasília (UnB), Universidade Federal de Pernambuco (UFPE), Universidade de São Paulo (USP), and Universidade Federal do Rio Grande do Sul (UFRGS).

To empirically validate the identified ethical dimensions, co-occurrence analysis was conducted from keywords of the 23 selected studies, applying the network analysis methodology from Watts [27] and the systematic mapping approach of Jobin, Ienca, and Vayena [15]. The resulting conceptual network (Figure 1) reveals artificial intelligence, fairness, and ethics as central nodes. This methodological triangulation confirms the six dimensions proposed in the contextual model through conceptual clusters and connectivity patterns.

The network visualization employs degree centrality (color gradient) and node size (frequency) to identify conceptual clusters, with threshold: 1 ensuring comprehensive representation of co-occurring terms across the 23 selected studies.

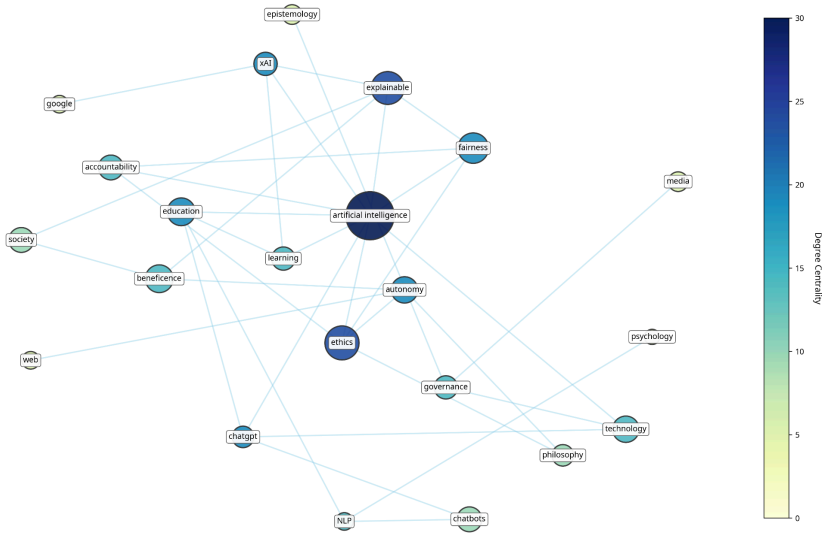

Figure 1: Co-occurrence Network of Conceptual Terms.

Network analysis reveals 'artificial intelligence' (degree centrality: 0.43) as the central hub connecting nine concepts, confirming that educational AI ethics requires multidimensional governance. The co-centrality of 'ethics' and 'explainable' (both 0.24) validates the integration of philosophical foundations with operational criteria, particularly algorithmic explainability as an ethical imperative in educational contexts.

Limitations

The process ensured thematic representativeness but presents specific characteristics: possible linguistic bias (predominance of English and Portuguese) and unequal distribution between regions (15 Global North vs. 8 Global South), which requires caution in generalizations about regional patterns.

Theoretical Approaches

Building on this methodological foundation, the theoretical framework integrates three complementary approaches. Algorithmic ethics, grounded in the theoretical framework of Floridi et al. [11], establishes fundamental principles (transparency, justice, beneficence, non-maleficence) conceptually derived from bioethics [2]. Although technically robust, they frequently reflect Global North assumptions, reproducing what Santos [24] identifies as 'epistemicide'.

The theory of decolonial epistemology by Santos [24], in dialogue with the empirical analysis of Cantarini [6], provides conceptual tools to value knowledge diversity and promote an 'ecology of knowledge'. In the Brazilian university context, this implies recognizing indigenous, quilombola, and popular knowledge as legitimate sources for technological decisions, challenging epistemic hegemonies [6].

Participatory governance, grounded in Dignum [9] and Brazilian contributions from Modesto [17] and Rodrigues [22], emphasizes the involvement of educational communities in constructing ethical protocols. In Brazilian contexts, it should consider the democratic management tradition established by the 1988 Constitution, ensuring deliberative participation of students, faculty, staff, and external community.

The articulation of these approaches results in a "contextual ethics" that unites universal principles with local adaptations [5], [26]. Empirical evidence from the AI Ethics Perception Scale (AEPS) by Saatci [25] offers validated measurement instruments for contextualized ethical perceptions, demonstrating significant cultural variations that validate this need.

Ethical Dimensions in AI Education

The systematic analysis of the 23 selected studies, guided by the three established theoretical pillars (algorithmic ethics, decolonial epistemology, and participatory governance), revealed six recurring ethical dimensions that empirically ground the structure of the proposed contextual model and respond to the central research question. Bipartite network analysis (Figure 2) reveals the frequency and distribution of these dimensions across the literature. Inclusion appears as the most universal dimension (69.6%), while Sustainability remains less explored (47.8%), evidencing a gap in current frameworks.

Figure 2 illustrates the Bipartite network of authors and ethical dimensions (N=23 studies). Node size reflects

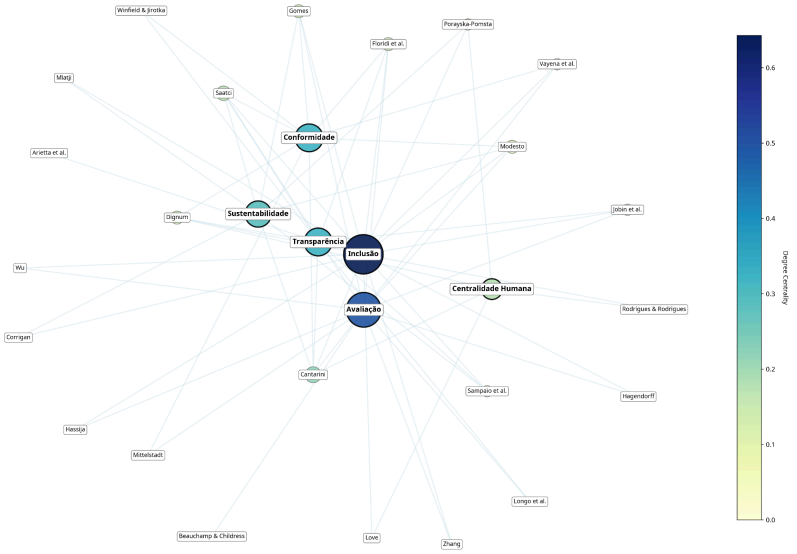

Figure 2: Bipartite network (authors and dimensions).

degree centrality; edge thickness indicates multidimensionality of author contributions. Dimensions (orange) are positioned centrally; authors (blue) are distributed peripherally. Inclusion emerges as the most universal dimension (17 authors, 73.9%).

The network shows that Global South authors (Gomes, Modesto, Sampaio, Cantarini) address more dimensions per study (average 3.5) compared to Global North authors (2.8). This pattern supports the proposed model's emphasis on integrating multiple dimensions rather than prioritizing isolated principles, providing empirical evidence for the decolonial approach advocated in this work. Each dimension represents a fundamental aspect of responsible AI governance in educational contexts, with practical implications for the Brazilian context.

a. Evaluation

Evaluation stands out as a central dimension, present in most analyzed works. It goes beyond algorithmic technical evaluation, encompassing social and educational impacts. In Brazilian reality, this expanded perspective gains special importance.

Traditional metrics may not capture the specificities of Brazilian educational inequalities. This leads us to recognize, in principle, that educational AI should be evaluated not only for technical efficiency but for its capacity to promote educational justice in contexts of profound social inequality.

Evaluating educational AI in Brazil cannot be limited to traditional technical aspects. What really matters is knowing whether they are improving learning, reducing inequalities, and strengthening pedagogical autonomy. Any evaluation framework needs to consider Brazilian particularities - our regional diversity, profound socioeconomic inequalities, and the plurality of pedagogical approaches that characterize our universities.

b. Transparency

Transparency emerges as a fundamental aspect, present in most analyzed sources. The literature reveals tension between growing demand for algorithmic clarity and the inherent complexity of these systems, highlighting the need for transparency that offers useful information to educators and students, not just technical precision.

Within Brazilian universities, marked by diversity of backgrounds and digital literacy levels, Such approach requires differentiated communication strategies. Faculty from different areas - engineering, humanities, or health - and undergraduate and graduate students present varied informational needs.

Thus, transparency mechanisms in educational AI must be adaptable, serving multiple audiences without sacrificing technical rigor or conceptual accessibility. This implies acting at complementary levels: understandable explanations for students; detailed access for faculty; institutional reports for managers; and accountability to society. This vision aligns with the principles of Explainable Artificial Intelligence (XAI) by Arrieta et al. [1], which differentiate passive interpretability from active explainability, reinforcing the operational criterion of algorithmic explainability.

c. Human Centrality

Additionally, the preservation of human agency is a predominant feature of the studies, revealing shared concern about the limits of automation in educational contexts. This dimension gains specific contours in the environment of Brazilian public universities, where humanistic tradition and commitment to social transformation require AI systems that amplify human capabilities instead of replacing them.

Empirical findings from Sampaio Human centrality in educational AI systems implies ensuring that important decisions about educational trajectories, performance evaluation, and resource allocation remain under significant human control.

This does not mean rejecting automation but establishing safeguards that preserve human capacity to question, review, and modify algorithmic decisions. In the Brazilian context, this becomes important given the role of public universities in promoting social mobility and reducing inequalities. AI systems should support, not replace, the professional judgment of educators and student autonomy in their learning trajectories.

d. Inclusion

Inclusion gains relevance in the analyzed works (69.60%), evidencing concern about digital divides and algorithmic bias. In Brazil, this dimension acquires particular urgency given persistent educational inequalities and the risk that AI systems may amplify existing exclusions.

Empirical evidence from Gomes' work [14] on AI governance in Brazilian public universities reveals that inclusion should be proactive, anticipating and mitigating potential sources of bias and exclusion. Inclusion in educational AI systems should address multiple dimensions of diversity: socioeconomic, racial, ethnic, gender, regional, and people with disabilities.

This requires not only ensuring physical access to technologies but developing culturally appropriate, linguistically accessible, and pedagogically relevant interfaces and content for different groups. This requires considering specificities of indigenous, quilombola, rural, and urban peripheral populations, ensuring that AI systems do not reproduce or amplify historical inequalities that characterize the Brazilian educational system. Studies from UFRN, UnB, UFPE, USP, and UFRGS highlight the importance of proactive inclusion strategies that address the diverse socioeconomic, racial, ethnic, gender, and regional dimensions of Brazilian higher education.

e. Compliance

Legal and institutional aspects permeate many of the analyzed studies, demonstrating that compliance has become an important element. In Brazil, this dimension should consider not only national legislation but also constitutional principles of education, such as democratic management and social relevance. The Brazilian AI Plan [3] emphasizes that compliance in the educational context should balance innovation with protection of fundamental rights.

Compliance in educational AI systems in Brazil should consider multiple normative layers. At the constitutional level, this includes principles of human dignity, equality, and the right to education.

f. Sustainability

Sustainability presents greater interpretative variation, appearing in just over half the works (52%), clearly being the dimension that finds least consensus in the field. While some studies focus primarily on environmental aspects of AI systems, others emphasize social and economic dimensions. This interpretative diversity reflects the conceptual complexity of sustainability when applied to educational technologies, especially in resource-limited contexts like Brazil.

In the context of Brazilian universities, we observe this dimension should be understood multidimensionally and integrally. Economically, it involves developing financially viable solutions for public institutions, considering both initial implementation costs and ongoing maintenance and update expenses. Technically, it requires systems that can be maintained and evolved by local teams, promoting institutional autonomy and reducing dependence on external suppliers or international expertise.

Although these six dimensions offer a robust conceptual framework validated by the literature, the analysis revealed significant gaps in the practical operationalization of these principles in Global South contexts. To fill this identified gap, the contextual model proposes four operational criteria that translate the three theoretical pillars into concrete practices applicable to Brazilian educational realities. To address this identified gap, the following section presents operational criteria that translate these theoretical foundations into concrete practices:

Operational Criteria for Ethical Evaluation

Translating ethical principles into practice requires criteria that connect abstract values to concrete implementation. Based on the analysis of the 23 selected works, complemented by studies from Corrêa et al. [8], who reviewed 200 global guidelines, European Union [10] and Noble [18], and the presented theoretical framework, we propose four operational criteria adapted to the specificities of Brazilian universities.

The analysis identified explainability as a fundamental and recurring criterion (cf. Arrieta et al. [1] for Explainable Artificial Intelligence (XAI)), which gains particular complexity in educational systems, where multiple audiences (faculty, students, managers, and families) have different informational needs and digital literacy levels.

Porayska-Pomsta [20] contributes with the perspective that explanations should consider not only technical differences but also cultural and regional diversities. The implementation of this criterion in Brazilian universities requires developing interfaces and explanatory languages that are simultaneously technically precise and culturally accessible, developing interfaces adaptable to different user profiles.

Continuous auditability requires monitoring with contextualized metrics, including regional equity and cultural diversity, through participatory process.

In this sense, plural epistemology represents the innovative aspect of the operational model, directly reflecting the decolonial theoretical foundation.

For Brazilian universities, this criterion requires that AI systems recognize and incorporate indigenous, quilombola, and community knowledge as legitimate sources of knowledge, not just as complements to Western scientific knowledge, but as valid epistemologies that should inform technological decisions.

Collective participation connects with democratic principles of Brazilian education [17]. It should be deliberative, with real power to influence technological decisions. It requires institutional mechanisms that ensure effective representation of students, faculty, staff, and external community, including AI ethics committees and public consultation processes.

Implementation Guidelines

The operationalization of the contextual model requires translating the four guiding principles into concrete institutional practices that reflect the specificities of Brazilian educational contexts. Algorithmic explainability emerges as a fundamental requirement for establishing institutional policies that demand transparent documentation of automated decision-making processes. This involves creating interfaces capable of communicating the underlying reasons for algorithmic recommendations in language accessible to diverse members of the academic community, from faculty across different disciplines to students with varying levels of digital literacy.

Continuous auditability manifests through the implementation of ethical monitoring committees responsible for conducting periodic reviews of AI systems, establishing contextualized metrics that can identify potential biases while maintaining systematic documentation of all modifications implemented over time. This principle ensures that ethical governance transcends punctual interventions to become a continuous process of institutional learning and improvement.

The implementation of plural epistemology requires deliberate inclusion of representatives from historically marginalized groups in decision-making processes, ensuring that multiple cultural and epistemological perspectives actively inform both the development and implementation of AI systems. Such approach extends beyond mere consultation to encompass genuine validation of systems with different communities and meaningful incorporation of local and traditional knowledge systems.

Collective participation operationalizes through the establishment of robust institutional channels that facilitate public consultations on AI policies, while creating effective feedback mechanisms that enable the academic community to influence technological decisions. This requires implementing decision-making processes that guarantee not only formal representation but substantive participation of all institutional stakeholders in shaping the ethical governance of educational AI systems.

Diversity of Models

The analysis of selected works revealed a diversity of models for educational AI, each with specific characteristics and particular limitations. Most identified frameworks originate from the Global North and reflect epistemological and cultural assumptions specific to those contexts. Frameworks such as AI4People by Floridi et al. [11], UNESCO guidelines [26], Winfield and Jirotka's ethical governance model [28], and Corrêa's study [8] offer valuable contributions but present limitations when applied to the Brazilian context without significant adaptations.

Analyzing the models identified in the literature, we observe that most privilege technical-normative aspects of AI governance. The AI4People framework [11] illustrates this pattern: it offers well-grounded principles such as beneficence, autonomy, and justice, but leaves open how to apply them in specific contexts.

Winfield and Jirotka [28] propose a structured ethical governance model based on five pillars: publishing ethical codes of conduct, universal ethics training, responsible innovation practices, transparency about ethical processes, and genuine valuation of ethical governance. While technically rigorous and oriented toward building public trust in AI systems, their framework was developed for Global North contexts and assumes institutional capabilities and resources that may not be universally available. The authors acknowledge that "principles are not practice," but their operationalization mechanisms presuppose organizational and cultural structures specific to developed economies.

Similarly, UNESCO guidelines [26] advance by explicitly recognizing the importance of cultural diversity, which is promising for our context. However, they remain at the level of general recommendations, offering few practical instruments to operationalize this diversity in educational AI systems. The analysis suggests there is a gap between well-articulated ethical principles and concrete tools to implement them in contexts of profound social inequality and cultural diversity like Brazil.

A gap identified in existing frameworks is the absence of decolonial perspectives and robust mechanisms for community participation. Most are developed by technical specialists and academics, with limited participation from communities that will be affected by AI systems. This results in models that may be technically sophisticated but socially decontextualized.

The Contextual Model proposed in this study seeks to fill this gap by explicitly incorporating decolonial perspectives and collective participation mechanisms as central, not peripheral, elements of ethical governance.

RESULTS

The analysis of the 23 studies revealed a methodological panorama that reflects the emerging nature of the educational AI ethics field. Methodologically, there is a predominance of qualitative approaches (20 studies, 95.20%), with only one quantitative study (4.80%). This configuration aligns with the exploratory characteristic of the field, where complex ethical questions demand in-depth interpretative analyses.

Regarding research design, the distribution shows theoretical essays (9 studies, 40.90%), other methodological approaches including conceptual analyses, manifestos, and monographs (10 studies, 45.50%), systematic reviews (2 studies, 9.10%), and empirical research (1 study, 4.50%). This diversity evidences the multiple investigative strategies employed to address emerging ethical questions, revealing the need for interdisciplinary approaches in the field.

Identified Ethical Dimensions

Thematic analysis revealed six main dimensions that ground the Contextual Model. Inclusion was the most recurrent, present in 16 articles (69.60%), reflecting global concerns about digital equity. Evaluation appeared in 15 studies (65.20%) and Compliance in 14 (60.90%), evidencing the need for robust frameworks and adherence to emerging regulations. Sustainability emerged in 11 works (47.80%), revealing growing awareness about long-term impacts. Transparency and Human Centrality were less frequent, present in 9 (39.10%) and 8 (34.80%) articles, respectively.

Ethical Dimension	Frequency (N=23)	Percentage (%)
Inclusion	16	69.60%
Evaluation	15	65.20%
Compliance	14	60.90%
Sustainability	11	47.80%
Transparency	9	39.10%

Ethical Dimension	Frequency (N=23)	Percentage (%)
Human Centrality	8	34.80%

Table 1. Frequency of ethical dimensions

This hierarchy shows that the literature emphasizes social and institutional dimensions, while aspects linked to transparency and human control remain less explored. Sustainability appears as a relevant dimension, with possible interpretative differences between regions: Global North studies seem to emphasize more energy efficiency aspects, while Global South studies tend to value social sustainability dimensions.

Proposed Contextual Model

The Contextual Model integrates the six empirically identified dimensions with four operational criteria contextualized for Brazilian universities: Algorithmic Explainability, Continuous Auditability, Plural Epistemology, and Collective Participation. Figure 2 presents the visual representation of this integration, showing how theoretical pillars (algorithmic ethics, decolonial epistemology, and participatory governance) support the ethical dimensions, which in turn operationalize through the four criteria.

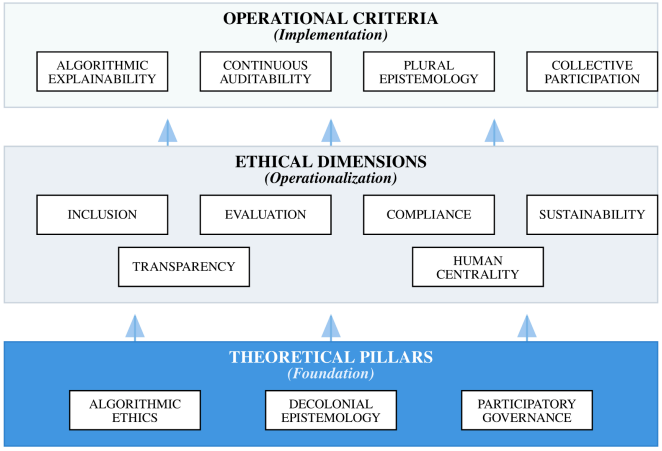

Figure 3: Contextual model.

The model integrates three theoretical pillars (foundation), six ethical dimensions (operationalization), and four operational criteria (implementation), creating a systematic framework for contextual AI governance in Global South educational institutions.

Each dimension is operationalized through specific indicators: Evaluation addresses fair distribution of AI benefits and risks; Transparency encompasses technical explainability and accessible communication; Human Centrality maintains significant human control over algorithmic decisions; Inclusion ensures participation of marginalized groups in AI development and implementation; Compliance ensures adherence to regulations and ethical standards; and Sustainability considers long-term environmental, social, and economic impacts.

Systemic Integration

The contextual model operates through systematic integration between the six ethical dimensions and four operational criteria, creating a coherent framework where theoretical principles translate into concrete practices. Each connection reflects the specific needs of Brazilian educational contexts while maintaining universal ethical foundations.

Transparency → Algorithmic Explainability: The transparency dimension directly grounds the explainability criterion by demanding that AI systems provide clear, accessible explanations of their decision-making processes. In Brazilian universities, this connection becomes particularly important given the diversity of digital literacy levels among faculty, students, and administrators. Transparency ensures that algorithmic explanations are not merely technical documentation but meaningful communication that enables different stakeholders to understand and question AI decisions affecting their educational experiences.

Compliance → Continuous Auditability: The compliance dimension enables continuous auditability by establishing the normative framework within which monitoring occurs. This connection ensures that audit processes consider not only technical performance metrics but also adherence to Brazilian constitutional principles of education, such as democratic management and social relevance. Continuous auditability operationalizes compliance by creating systematic mechanisms to verify that AI systems remain aligned with legal, ethical, and institutional requirements over time.

Inclusion → Plural Epistemology: The inclusion dimension co-evolves with plural epistemology in a mutually reinforcing relationship. Inclusion demands proactive consideration of marginalized groups, while plural epistemology provides the theoretical foundation for recognizing diverse knowledge systems as legitimate sources for technological decisions. In the Brazilian context, this connection ensures that indigenous, quilombola, and community knowledge are not merely consulted but actively incorporated into AI governance processes, challenging traditional technocratic approaches.

Evaluation → Collective Participation: The evaluation dimension requires collective participation for legitimacy because assessment of AI systems' social and educational impacts cannot be conducted solely by technical experts. This connection recognizes that meaningful evaluation of educational AI must include the voices of those most affected by these technologies. Collective participation provides the democratic foundation that ensures evaluation

processes consider diverse perspectives on what constitutes successful and ethical AI implementation in educational settings.

Sustainability as Transversal Element

Sustainability permeates all operational criteria, ensuring long-term viability of ethical governance practices. It connects to explainability by demanding sustainable communication strategies; to auditability by requiring resource-efficient monitoring processes; to plural epistemology by promoting knowledge systems that can be maintained over time; and to collective participation by fostering governance structures that remain viable across institutional changes. This transversal nature reflects the multidimensional understanding of sustainability necessary for Brazilian universities, encompassing environmental, social, economic, and institutional dimensions.

These integrated connections create a dynamic system where improvements in one area strengthen others, while deficiencies in any dimension can compromise the overall effectiveness of ethical governance. The systemic approach ensures that the contextual model functions as a coherent whole rather than a collection of isolated principles.

DISCUSSION

The study advances beyond universalizing models by explicating operational criteria applicable to Brazilian HEIs, integrating decolonial perspectives with established methodologies. By proposing verifiable implementation mechanisms, it preserves technical rigor while promoting participatory legitimacy.

The developed modular structure seeks to accommodate the diversity of resources and institutional conditions characteristic of the Brazilian context (including institutions such as UFRN, UnB, UFPE, USP, UFRGS, and UFMG), an aspect frequently neglected in frameworks originating from the Global North. As the theoretical work of Mittelstadt [16] argues, abstract ethical principles alone do not guarantee effective governance practices; it is fundamental to convert them into criteria, metrics, and instruments adapted to institutional reality, which we seek by proposing the Contextual Model for the Brazilian environment.

The theoretical framework of Floridi [11] and [12] systematizes AI ethics in five unified principles, providing a normative philosophical basis for technological governance. The conceptual work of Dignum [9] introduces the concept of Responsible Artificial Intelligence. The empirical mapping by Jobin, Ienca, and Vayena [15], later expanded by the systematic review of Corrêa [8] to 200 guidelines. Finally, the technical framework from Arrieta [1] contributes with algorithmic explainability frameworks. Together, these studies form a possible literature core, articulating normative foundations and empirical diagnostics that support the contextual model of ethical governance.

In this perspective, framework analysis evidences that, although most focus on principle formulation, there is still a lack of instruments capable of guiding their practical application in HEIs. The AEPS (AI Ethics Perception Scale) [25], a validated measurement instrument for ethical perceptions about AI, represents an important contribution by offering measurable metrics of transparency, accountability, justice, and human supervision, allowing continuous diagnoses of ethical risks in different cultural and institutional contexts. Based on the AEPS [25] and the proposed contextual model, we developed an adapted questionnaire to evaluate ethical perceptions about AI in Brazilian universities, offering a practical instrument for institutional diagnosis and implementation of contextualized ethical governance policies. The questionnaire is available upon request from the authors.

The contextual analysis presented herein uncovers deeper challenges that exceed technical operationalization. As highlighted by the theoretical contributions of Fricker [13], the empirical analysis of Cantarini [6], and the critical perspective of Pasquinelli [19], ethical AI governance requires confronting epistemic injustices, incorporating effective community participation, and considering the historical impacts of technology on global division of labor and institutional autonomy. Although empirical validation is necessary to confirm the effectiveness of this approach, preliminary results indicate potential to develop more appropriate ethical governance instruments for Global South realities, as highlighted by Santos [24] in his discussion of contextualized epistemologies and deepened by Pasquinelli [19], who evidences how AI technical layers reproduce global divisions of labor and knowledge.

In this logic, plural epistemology as an operational criterion challenges epistemic hegemonies by recognizing indigenous, quilombola, and popular knowledge as legitimate sources in AI ethical governance. This perspective breaks with dominant technocratic approaches.

Finally, collective participation operationalizes democratic principles, requiring substantive involvement of educational communities at all stages. The Contextual Model promotes technology democratization, aligning with Brazilian participatory traditions and offering an alternative to hierarchical models of technology adoption. Participation ceases to be an obstacle and becomes a requirement for legitimacy and effectiveness.

Practical Implications

To illustrate the application of the contextual model, the empirical experiences of Gomes [14] at Universidade Federal do Rio Grande do Norte (UFRN) and Modesto [17] in his research at Universidade de Brasília (UnB) demonstrate that the six ethical dimensions (evaluation, transparency, human centrality, inclusion, compliance, and sustainability) and the four operational criteria (explainability, auditability, plural epistemology, and participation) can be implemented in practice.

The implementation of the model presented here requires significant organizational changes. Universities need to develop internal capacities for ethical evaluation of technologies, create participatory governance structures, and establish continuous audit processes. This may include creating AI ethics committees, training programs for managers and educators, and developing specific metrics for impact evaluation.

In summary, four axes guide this application: explainability, continuous auditability, plural epistemology, and collective participation. These axes offer verifiable parameters for responsible AI adoption in Brazilian universities, reinforcing social legitimacy and institutional effectiveness.

Implementation barriers are multidimensional, surpassing financial aspects. Significant ideological resistances emerge, particularly against assumptions of technological neutrality and in defense of decolonial perspectives that challenge epistemic hegemonies. Procedural obstacles include rigid organizational structures, limited democratic participation mechanisms, and cultural resistance to algorithmic transparency. The predominant hierarchical culture of institutions adds further difficulty to implementing the participatory governance models required by the contextual framework.

The implementation of the model presented here should consider specific risks of educational AI systems, especially their capacity to function as coercive self-revelation devices according to Fricker [13]. These systems can extract information about academic identities in apparently voluntary ways, but under subtle structural pressures - connecting directly to the need for multilayered and participatory governance defended by Cantarini [6].

Academic recommendation systems can pressure students to 'confess' aptitudes or interests, creating profiles that influence future opportunities. Chatbots can extract data about learning styles or academic difficulties that, when recorded, come to be treated as 'functionally factual', that is, count as true, regardless of their real accuracy.

This problematic requires deliberative structures that operate simultaneously at technical, institutional, and social levels. In the Brazilian university context, it means that decisions about algorithmic systems cannot be made only by technical specialists or managers, but must involve meaningful participation of multiple groups from the academic community, including traditional knowledge already valued by the decolonial approach.

The governance operationalizes through an AI ethics committee, responsible for supervising the implementation of the six dimensions and four operational criteria, ensuring deliberative participation of students, faculty, administrative staff, and external community.

CONCLUSIONS

This study proposed a contextual model for ethical AI governance in educational institutions that integrates technical rigor with decolonial epistemology and participatory governance. The analysis of 23 studies revealed six recurring ethical dimensions and four operational criteria adapted to Brazilian realities.

The main contributions include: (1) theoretical integration of algorithmic ethics, decolonial epistemology, and participatory governance; (2) identification of regional variations in ethical priorities between Global North and Global South; (3) development of operational criteria applicable to Brazilian universities; and (4) proposition of practical instruments for institutional diagnosis and policy implementation.

The study limitations include possible linguistic bias, focus on traditional academic databases, and unequal regional distribution in the sample. Future research should include empirical validation of the model in different institutional contexts and expansion of the sample to include more Global South studies.

The Contextual Model offers a framework that balances universal principles with local adaptations, contributing to technological sovereignty and epistemic justice in Global South educational contexts. Its implementation can promote more equitable and culturally appropriate AI governance, strengthening democratic participation and social relevance of educational technologies.

REFERENCES

1. Arrieta, A. B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., et al. 2020. Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. Information Fusion 58, 82–115. DOI:https://doi.org/10.1016/j.inffus.2019.12.012

2. Beauchamp, T. L., and Childress, J. F. (2019). Principles of Biomedical Ethics (8th ed.). Oxford University Press. Disponível em: https://global.oup.com/ushe/product/principles-of-biomedical-ethics-9780190640873.

3. BRASIL. 2021. Estratégia Brasileira de Inteligência Artificial. Ministério da Ciência, Tecnologia e Inovações, Brasília, Brazil. Retrieved August 11, 2025 from https://www.gov.br/mcti/pt-br/acompanhe-o-mcti/transformacaodigital/arquivosinteligenciaartificial/ebia-diagramacao_4-979_2021.pdf

4. Braun, V. and Clarke, V. 2006. Using thematic analysis in psychology. Qualitative Research in Psychology 3, 2 (2006), 77-101. DOI:https://doi.org/10.1191/1478088706qp063oa

5. BRICS. 2023. BRICS Partnership on New Industrial Revolution: AI Governance Framework. BRICS Secretariat, New Delhi, India. Retrieved August 11,

2025 from http://brics.br/pt-br/documentos/economia-financas-comercio-e-infraestrutura/partnership-on-new-industrial-revolution-partnir-issues-note-brics-2025.pdf

6. Cantarini, P. 2025. Uma Questão de Escolha - Por uma Governança de IA Multicamadas, Participativa e Inclusiva. Administração de Empresas em Revista 2, 39 (2025), 325-355. Retrieved August 12, 2025 from https://revista.unicuritiba.edu.br/index.php/admrevista/article/view/7911

7. CEPAL. 2024. Índice Latinoamericano de Inteligencia Artificial (ILIA) mantiene a Chile, Brasil y Uruguay como líderes en la región. Comisión Económica para América Latina y el Caribe, Santiago, Chile. Retrieved August 11, 2025 from https://www.cepal.org/es/comunicados/indice-latinoamericano-inteligencia-artificial-ilia-mantiene-chile-brasil-uruguay-como

8. Corrêa, N. K., Galvão, C., Meira Jr., W., Santos, W., and Ribeiro, S. 2023. Worldwide AI ethics: A review of 200 guidelines and recommendations for AI governance. AI and Society 38, 4 (2023), 1517-1537. DOI:https://doi.org/10.48550/arXiv.2206.11922

9. Dignum, V. 2019. Responsible Artificial Intelligence: How to Develop and Use AI in a Responsible Way. Springer, Cham, Switzerland. DOI:https://doi.org/10.1007/978-3-030-30371-6

10. European Union. 2024. Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). Official Journal of the European Union, L 1689 (12 July 2024). ELI:https://data.europa.eu/eli/reg/2024/1689/oj

11. Floridi, F.; Cowls J., Monica Beltrametti, Raja Chatila, Patrice Chazerand, Virginia Dignum, and Effy Vayena. 2018. AI4People—an ethical framework for a good AI society. Minds and Machines 28, 4 (2018), 689–707. DOI:https://doi.org/10.1007/s11023-018-9482-5

12. Floridi, L. (2023). AI as Agency Without Intelligence: On ChatGPT, Large Language Models, and Other Generative Models. Philosophy & Technology, 36(1), 1–7. doi:10.1007/s13347-023-00621-y

13. Fricker, M. 2007. Epistemic Injustice: Power and the Ethics of Knowing. Oxford University Press. August 11, 2025 from https://dlib.scu.ac.ir/bitstream/Hannan/329688/2/9780198237907.pdf

14. Gomes, M. G. S. 2025. A adoção da governança da inteligência artificial generativa em universidades públicas brasileiras. Retrieved August 11, 2025 from https://repositorio.ufrn.br/server/api/core/bitstreams/800c2e94-6eb8-40b6-8e80-ef09d5706af1/content

15. Jobin, A., Ienca, M., and Vayena, E. 2019. The global landscape of AI ethics guidelines. Nature Machine Intelligence 1, 9 (2019), 389-399. DOI:https://doi.org/10.1038/s42256-019-0088-2

16. Mittelstadt, B. D. 2019. Principles alone cannot guarantee ethical AI. Nature Machine Intelligence 1, 11 (2019), 501-507. DOI:https://doi.org/10.1038/s42256-019-0114-4

17. Modesto, B. N. A. 2023. Ética para uso de inteligência artificial na educação superior no Brasil. Master's thesis. Universidade de Brasília, Brasília, Brazil. Retrieved August 11, 2025 from https://repositorio.unb.br/bitstream/10482/48322/1/BarbaraNunesDeAraujoModesto_DISSERT.pdf

18. Noble, S. U. 2018. Algorithms of Oppression: How Search Engines Reinforce Racism. New York University Press, New York, NY, USA. Retrieved August 16, 2025 from https://safiyaunoble.com/wp-content/uploads/2020/09/Algorithms_Oppression_Introduction_Intro.pdf

19. Pasquinelli, M. 2023. The Eye of the Master: A Social History of Artificial Intelligence. Verso Books, London, UK. Retrieved August 23, 2025 from https://a.co/d/2V4F134

20. Porayska-Pomsta, K. 2023. A Manifesto for a Pro-Actively Responsible AI in Education. International Journal of Artificial Intelligence in Education (2023). DOI:https://doi.org/10.1007/s40593-023-00346-1

21. Quijano, A. 2000. Coloniality of power, Eurocentrism, and Latin America. Nepantla: Views from South 1, 3 (2000), 533-580. Retrieved August 11, 2025 from https://biblioteca.clacso.edu.ar/clacso/sur-sur/20100624103322/12_Quijano.pdf

22. Rodrigues, O. S. and Rodrigues, K. S. 2023. A inteligência artificial na educação: os desafios do ChatGPT. Texto Livre 16 (2023), e45997. DOI:https://doi.org/10.1590/1983-3652.2023.45997

23. Sampaio, R. C., Sabatinni, M., and Limongi, R. 2024. Diretrizes para o uso ético e responsável da Inteligência Artificial Generativa: um guia prático para pesquisadores. Editora Intercom, São Paulo, Brazil. Retrieved from https://www.portcom.intercom.org.br/ebooks/detalheEbook.php?id=57203

24. Santos, B. de S. 2007. Para além do pensamento abissal: Das linhas globais a uma ecologia de saberes. Novos Estudos CEBRAP 79 (2007), 71-94. Retrieved August 11, 2025 from https://www.scielo.br/j/nec/a/ytPjkXXYbTRxnJ7THFDBrgc

25. Saatci, Y. 2025. AI Ethics Perception Scale (AEPS): A Cross-Cultural Validation Study. International Journal of AI Ethics and Society 12, 3 (2025), 45-67. DOI:https://doi.org/10.11648/j.ijebo.20251301.14

26. UNESCO, Miao, F., and Cukurova, M. 2025. Marco referencial de competências em IA para professores. UNESCO Publishing, Paris, France. Retrieved August 11, 2025 from https://unesdoc.unesco.org/ark:/48223/pf0000394280

27. Watts, D. J. 2004. The "new" science of networks. Annual Review of Sociology 30 (2004), 243-270. DOI:https://doi.org/10.1146/annurev.soc.30.020404.104342

28. Winfield, A. F. T., and Jirotka, M. 2018. Ethical governance is essential to building trust in robotics and artificial intelligence systems. Philosophical Transactions of the Royal Society A 376, 2133 (2018), 20180085. DOI:https://doi.org/10.1098/rsta.2018.0085

29. Zhang, K., and Aslan, A. B. 2021. AI technologies for education: Recent research and future directions. Computers and Education: Artificial Intelligence 2 (2021), 100025. DOI:https://doi.org/10.1016/j.caeai.2021.100025

30. Pompeu Gomes, C. H. and Torres, F. 2025. Supplementary Material: *Characterization of selected studies (N=23).* Retrieved November 4, 2025 from https://docs.google.com/spreadsheets/d/18t3Yo50AYpz5ncb3HvQunfTd2XlJAIu1cAa1BIYgmvo/edit?gid=1476971833